(51) International Patent Classification 6 : C12Q 1/68, C12N 15/10, 15/85, 15/62	A1	(11) International Publication Number: WO 99/43848 (43) International Publication Date: 2 September 1999 (02.09.99)

(21) International Application Number: PCT/CA99/00173

(22) International Filing Date: 25 February 1999 (25.02.99)

(30) Priority Data:
2,224,475   25 February 1998 (25.02.98)   CA

(71) Applicant (for all designated States except US): THE UNIVERSITY OF BRITISH COLUMBIA [CA/CA]; The UBC University–Industry Liaison Office, IRC Room 331, 2194 Health Sciences Mall, Vancouver, British Columbia V6T 1Z3 (CA).

(72) Inventors; and
(75) Inventors/Applicants (for US only): ONG, Christopher, J. [CA/CA]; 101 – 688 West 12th Avenue, Vancouver, British Columbia V5Z 1M8 (CA). JIRIK, Frank, R. [CA/CA]; 103 – 146 West 13th Avenue, Vancouver, British Columbia V5Y 1V7 (CA).

(74) Agents: ROBINSON, J., Christopher et al.; Smart & Biggar, Box 11560, Suite 2200, 650 West Georgia Street, Vancouver, British Columbia V6B 4N8 (CA).

(81) Designated States: CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
   With international search report.
   Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: PROTEIN INTERACTION AND TRANSCRIPTION FACTOR TRAP

(57) Abstract

Methods are provided which make use of a combination of gene trap and two–hybrid methodologies for the identification and characterization of unknown genes according to protein–protein interactions of the gene product or for the identification and characterization of unknown genes encoding transcriptional activator domains (AD). Interaction of an exon–encoded protein domain with a known protein, or functioning of the exon–encoded domain as an AD, is detected by reconstituting the activity of a transcriptional activator. Suitable gene trap vectors are also provided.

PROTEIN INTERACTION AND TRANSCRIPTION FACTOR TRAP

Field of Invention

5   This invention relates to the use of gene trapping
methods for the identification of genes and two-hybrid
methodology for the identification of protein-protein
interactions.

Background of the Invention

       Virtually all cellular responses, including growth and
differentiation,   are   stringently   controlled   by
physiological signals in the form of growth factors,
15   hormones, nutrients, and contact with neighbouring cells.
These various signals are processed and interpreted by
signal transduction mechanisms which ultimately induce the
cell to mount an appropriate response. Signalling pathways
stimulated by physiological signals involves a network of
20   specific protein-protein interactions which function to
transmit the signal to downstream effector molecules that
execute the response. Thus, specific interactions between
proteins are critical for signal transduction mechanisms as
well as regulation of cellular architecture and responses
25   to   physiological   signals.   Given   that   specific
protein-protein interactions are involved in execution of
virtually   all   cellular   functions,   technologies   which
simplify and facilitate detection and analysis of specific
protein-protein interactions will be valuable for the
30   discovery, design and testing of drugs that target highly
specific biological processes.

       Eukaryotic gene expression is regulated by a class of
proteins variously known as transcriptional activators, or
35   enhancer binding proteins and are referred to herein as
"transcriptional regulatory proteins". These molecules,
bind to specific sequences on DNA within the promoters of
genes they regulate, and function by recruiting the general
transcriptional   initiation   complex   to   the   site   where

transcription of DNA into messenger RNA (mRNA) begins. The general eukaryotic transcriptional initiation complex may consist of two large protein complexes represented by transcription factor IID (TFIID), which contains the

5   TATA-element binding protein that functions to position the general initiation complex at a precise location on the promoter, and the RNA polymerase II holoenzyme, which contains the catalytic function necessary to unwind the double stranded DNA and transcribe a copy of the DNA

10   template into mRNA. Known transcriptional activators are understood to function by forming direct protein-protein interactions with parts of TFIID and/or the RNA polymerase holoenzyme, and catalysing their assembly into an initiation complex at TATA-element of the promoter.

15

Transcriptional regulatory proteins typically possess two functional elements, a site-specific DNA-binding domain and a transcriptional activation domain which can interact with either TFIID or the RNA polymerase holoenzyme.

20   Eukaryotic transcriptional regulatory proteins are typified by the *Saccharomyces* yeast GAL4 protein, which was one of the first eukaryotic transcriptional activators on which these functional elements were characterized. GAL4 is responsible for regulation of genes which are necessary for

25   utilization of the six carbon sugar galactose. Galactose must be converted into glucose prior to catabolism; in *Saccharomyces* this process typically involves four reactions which are catalysed by five different enzymes. Each enzyme is encoded by a GAL gene (GAL 1, 2, 5, 7, and

30   10) which is regulated by the transactivator GAL4 in response to the presence of galactose. Each GAL gene has a *cis*-element within the promoter, termed the upstream activating sequence for galactose (UAS$_G$), which contains 17 base-pair sequences to which GAL4 specifically binds.

35   The GAL genes are repressed when galactose is absent, but are strongly and rapidly induced by the presence of

galactose. GAL4 is prevented from activating transcription when galactose is absent by a regulatory protein GAL80. GAL80 binds directly to GAL4 and likely functions preventing interaction between GAL4's activation domains and the general transcriptional initiation factors. When yeast are given galactose, transcription of the GAL genes is induced. Galactose causes a change in the interaction between GAL4 and GAL80 such that GAL4's activation domains become exposed to allow contact with the general transcription factors represented by TFIID and the RNA polymerase II holoenzyme and catalyse their assembly at the TATA-element which results in transcription of the GAL genes. The functional regions of GAL4 have been defined by a combination of biochemical and molecular genetic strategies. GAL4 binds as a dimer to its specific cis-element within the UAS_G of the GAL genes. The ability to form tight dimers and bind specifically to DNA is conferred by an N-terminal DNA-binding domain. This fragment of GAL4 (amino acids 1-147) can bind efficiently and specifically to DNA but cannot activate transcription. Two parts of the GAL4 protein are necessary for activation of transcription, called activating region 1 and activating region 2. The activating regions are thought to function by interacting with the general transcription factors. The large central portion of GAL4 between the two activating regions is required for inhibition of GAL4 in response to the presence of glucose. The C-terminal amino acids of GAL4 bind the negative regulatory protein GAL80; deletion of this segment causes constitutive induction of GAL transcription.

An important contribution towards development of two-hybrid methodology was the discovery that a transcriptional activator protein, the Herpes viral protein 16 (VP16), is indirectly recruited to DNA through interaction with sequence specific DNA binding protein.

VP16 activates transcription by forming a complex with the cellular proteins Oct-1 and HCF; the Oct-1/HCF/VP16 complex bind to enhancer elements of the Herpes immediate early genes. It was subsequently shown that the negative regulatory protein GAL80 could be converted into a GAL4-dependent transactivator by fusion of a short negatively-charged transcriptional activating sequence B17. The GAL80-B17 fusion protein, when co-expressed with GAL4, was found to cause activation of a GAL4-dependent reporter gene to a greater extent than GAL4 alone.

The standard two-hybrid assay relies upon the fact that many eukaryotic transcriptional regulatory systems consist of the separate domains discussed above: the DNA-binding domain (DNA-BD) that binds to a promoter or other *cis*-transcriptional regulatory element; and, the activation domain (AD) that directs RNA polymerase II to transcribe a gene downstream from the site on the DNA where the DNA-BD is bound. The DNA binding domain and the activation domain may be separate proteins but will function to activate transcription as long as the AD is in proximity to a DNA-BD bound to the transcriptional regulatory element. Where each of the AD and the DNA-BD is fused to members of a pair of interacting proteins, the AD will function via the link to the DNA-BD created by the interacting proteins. Thus, the two-hybrid assay may be used to investigate whether interaction occurs between two proteins (termed "bait" and "prey") expressed as fusion products with DNA-BD and AD peptides, respectively. A positive event is identified by activation of a reporter gene having an upstream promoter to which the DNA-BD binds.

The two-hybrid assay may be carried out in a variety of eukaryotic cells including yeast (see: Fields, S. and Song. O. 1989 *A Novel Genetic System to Detect Protein-Protein Interactions* Nature 340:245-247; and

Fields, S. 1993. *The Two-hybrid System to Detect Protein-Protein Interactions*. Methods: A Companion to Meth. Enzymol. 5:116-124.) and mammalian cells (see: Luo, Y. *et al.* 1997. *Mammalian Two-Hybrid System: A*
5   *Complementary Approach to the Yeast Two-Hybrid System*. Biotechnics 22:350-52; and, Feron, E.R. *et al.* 1992. *Karyoplasmic Interaction Selection Strategy: A General Strategy to Detect Protein-Protein Interactions in Mammalian Cells* Proc. Natl. Acad. Sci. U.S.A. 89:7958-62).
10 Commercial yeast and mammalian two-hybrid assay kits are available from Clontech Laboratories, Inc., 1020 East Meadow Circle, Palo Alto, California, 94303-4230, U.S.A.

A variant of the two-hybrid method, called the
15 "interaction trap" system, employs the principle of using separate fusions with DNA-binding and transactivation domains, except that the bait is fused to LexA, which is a sequence-specific DNA binding protein from E. coli, and an artificial transactivation domain known as B42 (31) is used
20 for the "prey" fusions. Interaction between the bait and prey fusions is detected by expressed of a LexA-responsive reporter gene.

A modification of the standard two-hybrid system known
25 as "Reverse Two-Hybrid" (Erickson *et al.* U.S. Pat No. 5,535,490; Vidal *et al.* International Application Number PCT/US96/04995) has been described which is intended for use in identifying specific inhibitors of a standard two-hybrid protein-protein interaction. The reverse two-
30 hybrid system operates by driving the expression of relay gene, such as the GAL80 gene, that encodes a protein that bind to and masks the activation domain of a transcriptional activator such as GAL4. Expression of the reporter gene is made dependent upon the functioning of the
35 activation domain of the transcriptional activator. Only when the level of the masking protein is reduced because a

compound interferes with the two-hybrid interaction will the activation domain of the transcriptional activator be unmasked and allowed to function.

5   Specific protein-protein interactions are the basis for many biological processes. Standard two-hybrid techniques make use of specialized cDNA expression libraries as a source of protein sequences used in screening for specific interactions between proteins (for
10   example in drug screening programs). However, cDNA expression libraries possess some intrinsic disadvantages. For example, cDNA libraries produce a bias toward cloning of highly expressed genes and rare gene transcripts are unlikely to be discovered. The source of the mRNA for the
15   generation of the cDNA library is critical since many tissue restricted genes and developmentally or temporally regulated genes are not represented by a particular cDNA library.

20   Gene trap vectors target the prevalent introns of the eukaryotic genome. These vectors may consist of either a splice-acceptor (SA) site upstream of a reporter sequence, or an unpaired splice-donor (SD) site downstream from a reporter sequence. Preferably, on the latter vector
25   comprising a SD, the reporter sequence is driven by an appropriate transcriptional regulatory element (eg. promoter). Integration of the above-described gene trap vectors into an intron results in production of m-RNA in which a transcript of the vector is joined to an
30   transcript of an adjacent exon. (see: Skarnes, W.C. et al. 1992. A Gene Trap Approach in Mouse Embryonic Stem Cells: The lacZ Reporter is Activated by Splicing, Reflex Endogenous Gene Expression and is Mutagenic in Mice. Genes Dev. 6:903-918; W.C. Skarnes 1993 The Identification of New
35   Genes: Gene Trapping in Transgenic Mice. Current Opinion in Biotechnology 4:684-89; and, United States Patent

No. 5,652,128 July 29, 1997.). A form of gene trapping
(termed "tagging") may also be accomplished by using a
vector comprising a peptide encoding segment and both an
upstream SA and a downstream SD (see United States Patent
5   No. 5,652,128 of Jarvik).

Features of gene trapping include:

(a)   random integration into the genome;

10

(b)   splice   acceptor   or   splice   donor   containing
vectors result in fusion of a transcript of a
reporter gene from the vector with endogenous
gene transcripts;

15

(c)   the full repertoire of genes are represented in
the genome without a bias towards highly
expressed genes;

20   (d)   gene trapping can provide information about
coding regions of most genes that is independent
of their transcription status; and

(e)   gene trapping is independent of the source of
25   mRNA (therefore, rare as well as tissue specific
genes and developmental temporally regulated
genes may be trapped).

A full strategy for genome-wide analysis as well as
30   for drug discovery and assessment, should include a
systematic strategy for identification and characterization
of gene products according to their protein-protein
interaction characteristics.

35

- 8 -

Summary of Invention

Gene trap methodologies provide a repertoire of
protein domains encoded by exon sequences found within the
5   genome. Two-hybrid techniques permit identification of
protein-protein interactions. This invention makes use of
a combination of gene trap and two-hybrid methodologies for
the identification and characterization of genes according
to protein-protein interactions of the gene product or for
10   the identification of genes encoding transcriptional
activator domains (AD). Interaction of an exon-encoded
protein domain with a given protein, or functioning of the
exon-encoded domain as an AD, is detected by reconstituting
the activity of a transcriptional activator.
15

This invention also provides gene trap vectors adapted
for use in a two-hybrid assay and methodologies for
identification of genes encoding proteins capable of
interacting with a selected protein. This invention also
20   provides gene trap vectors and methodologies for the
selective identification of genes encoding transcription
activator domains.

This invention provides a DNA construct comprising a
25   DNA sequence encoding a transcriptional regulatory protein
moiety selected from the group consisting of a DNA-BD and
a AD; and, a m-RNA splice site. The term "m-RNA" splice
site is defined herein as being a splice acceptor sequence
(SA), an unpaired splice donor sequence (SD).
30

This invention also provides a DNA construct
comprising a DNA sequence encoding a transcriptional
regulatory protein moiety selected from the group
consisting of a DNA-BD and an AD; and, a downstream SD.
35   This DNA construct preferably contains no nucleic acid
sequence which would encode a protein that will interact

with a test protein employed in this invention. Preferably, the only protein encoded by the construct or the portion of the construct between the 5' end of the sequence encoding the transcriptional regulatory protein moiety and the 3' end of the SD is the transcriptional regulatory protein moiety itself. Preferably, this construct will have a transcriptional regulatory element (eg. a promoter) operably linked to the sequence encoding the transcriptional regulatory protein moiety.

This invention also provides a DNA construct comprising a DNA sequence encoding a transcription regulatory protein moiety selected from the group consisting of a DNA-BD and an AD, together with an upstream SA and a downstream SD. Alternatively, this DNA construct may comprise an SA upstream of a transcriptional regulatory protein moiety selected from the group consisting of a DNA-BD and an AD; and, a downstream poly-adenylation signal. Preferably, these DNA constructs will not encode any protein which will interact with a test protein as used on this invention. Preferably, the only protein encoded by the construct or the portion of the construct between the SA and the SD or the SA and the poly-adenylation signal, will be the transcriptional regulatory protein moiety.

This invention also provides a method of making the DNA constructs of this invention comprising the step of joining a DNA sequence encoding a transcriptional regulatory protein moiety as defined above with one or both of a SA and a SD. Preferably, at least three such DNA constructs are made in three different reading frames. This invention also provides cells comprising the DNA constructs of this invention obtainable by the method of transforming eucaryotic cells with one or more DNA constructs of this invention.

This invention also provides kits that comprise the above-described DNA construct of this invention. The DNA constructs may be in the form of plasmids. The kits may also comprise host cells, two-hybrid vectors or reporter

5   gene constructs as described herein. The two-hybrid vectors of the kit may be plasmids constructed (eg. presence of suitable restriction sites) to permit insertion of a test protein sequence to be part of a two-hybrid vector as described herein. The kits may also

10   comprise materials and reagents useful for DNA insertions, reporter gene activity assays, or sequencing of inserts (eg. primers).

This invention also provides host cells whose genome

15   optionally comprises a reporter gene as described herein and wherein the cell expresses a two-hybrid vector as described herein. The two-hybrid vector may include a sequence encoding a test protein.

20   This invention also provides a method for detecting interaction between an endogenous protein of a cell and a test protein, wherein said cell contains a first DNA sequence encoding a reporter under transcriptional control of a transcriptional regulatory element, and a second DNA

25   sequence that is expressed by the cell and which encodes a first hybrid protein comprising:

     (a) a first transcriptional regulatory protein moiety
          selected from the group consisting of: a DNA-BD

30   that   recognizes   a   binding   site   on   the
          transcriptional regulatory element controlling
          the first DNA sequence and, a AD functional in
          the cell; and

35   (b) a test protein;

wherein the method comprises the steps of:

(a)   placing into the cell or an ancestor of the cell,
a DNA construct comprising one or more m-RNA
splice sites, and a third DNA sequence encoding
a second transcriptional regulatory protein
moiety which, when combined with the first
transcriptional regulatory protein moiety will
reconstitute a transcriptional regulatory protein
capable of binding to and activating the
transcriptional regulatory element controlling
transcription of the first DNA sequence; and,

(b)   determining whether the reporter is expressed by
the cell or a descendant of the cell, as an
indicator of expression of a second hybrid
protein comprising the second transcriptional
regulatory protein moiety and an endogenous
protein of the cell capable of interaction with
the test protein.

The DNA construct comprising a third DNA sequence
described in the method above may be selected from the
following group, in which it is preferable that the only
protein encoded by the construct or the portion of the
construct described above, be the transcriptional
regulatory moiety itself:

(I)   a gene trap vector comprising the third DNA
sequence to reconstitute a transcriptional
regulatory protein, followed by a SD; and
preferably, a transcriptional regulatory element
operably linked to the third DNA sequence;

(II)   a gene trap vector without a transcriptional
regulatory element and comprising a SA upstream

of the third DNA sequence to reconstitute a transcriptional regulatory protein; and preferably, the third DNA sequence is followed by a poly-adenylation signal; and

5

(III) a gene trap vector comprising the third DNA sequence to reconstitute a transcriptional regulatory protein, with an upstream SA and a downstream SD.

10

In embodiments of the method described above in which the second DNA sequence encodes a DNA-BD that recognizes a binding site on the transcriptional regulatory element controlling the reporter, the DNA construct comprising the third DNA sequence will encode an AD. Where the second DNA sequence encodes an AD, the DNA construct will comprise a DNA-BD capable of binding to the transcriptional regulatory element controlling the reporter. When the second nucleotide sequence is expressed in a cell in which the third DNA sequence is also expressed (resulting in a hybrid protein containing an endogenous portion that interacts with the test protein) reconstitution of the transcriptional regulatory protein occurs. Binding of the latter protein by means of the DNA-BD to the transcriptional regulatory element controlling the reporter results in expression of the reporter.

In the method described above, the third DNA sequence will preferably encode an AD, not a DNA-BD. This may minimize false positives resulting from reconstitution of a transcriptional regulatory protein when the third DNA sequence is expressed with an exon that encodes an endogenous protein that itself is capable of functioning as an AD.

35

This invention also provides a method for detecting an endogenous transcription activator domain (AD) of a cell, wherein the cell contains a first DNA sequence encoding a reporter under transcriptional control of a transcriptional regulatory element, wherein the method comprises the steps of:

(a)   placing into the cell or an ancestor of the cell, a DNA construct comprising a m-RNA splice site and a second DNA sequence encoding a DNA-BD that recognizes a binding site on the transcriptional regulatory element controlling transcription of the first DNA sequence; and

(b)   detecting expression of the reporter in the cell or a descendant of the cell, as an indicator of expression of a hybrid protein comprising the DNA-BD and an endogenous protein of the cell capable of functioning as an activator domain.

The DNA construct comprising a second DNA sequence as used in the above-described method for detecting an endogenous transcription activator domain may be selected from the following group in which it is preferred that the only protein encoded by the construct or the portion of the construct described above, be the transcriptional regulatory moiety itself:

(IV)   a gene trap vector comprising the second DNA sequence, followed by a SD; and preferably, a transcriptional regulatory element is operably linked to the second DNA sequence;

(V)   a gene trap vector without a transcriptional regulatory element and comprising a SA upstream of the second DNA sequence; and preferably, the

- 14 -

second DNA sequence is followed by a poly-adenylation signal; and

(VI) a gene trap vector comprising the second DNA
5   sequence, with an upstream SA and a downstream SD.

Detailed Description of the Invention

10   The terms "host cell" and "cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations
15   due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent or ancestor cell, but are still included within the scope of the terms as used herein. A cell as used in the method of this invention is a eukaryotic cell.
20

A "DNA construct" is a deoxynucleic acid (DNA) molecule, either single- or double-stranded, that has been modified through human intervention to contain segments of DNA combined and juxtaposed in an arrangement not existing
25   in nature.

A "reporter" as used herein, may refer to a polynucleotide sequence (structural sequence) encoding a reporter protein or the term may refer to the reporter
30   protein itself, depending upon the context.

The term "operably linked" is intended to mean that a DNA sequence is linked to a regulatory sequence in a manner which allows expression of the DNA sequence. Such a
35   regulatory sequence includes promoters, enhancers and other expression control elements.

- 15 -

The terms "polypeptide", "peptide" and "protein" as used herein refer to a polymer of amino acid residues.

The term "endogenous" refers to that which is produced
5   or arises from within a cell or organism.

The term "plasmid" refers to a circular, double stranded, extrachromosomal bacterial DNA into which additional DNA segments may be ligated and which replicates
10   automatically.   Methodologies   for   selection   and construction of vectors, plasmids and DNA constructs may be found, for example in: Molecular Cloning: A Laboratory Manual; (2d), Sambrook et al. 1989, Cold Spring Harbor Laboratory Press. Suitable host cells are discussed
15   further in Goeddel; "Gene Expression Technology" in: Methods in Enzymology 185, Academic Press, San Diego, California (1990).

In the present invention, DNA constructs are
20   introduced into a host cell and expressed in the host cell in sufficient quantities for a reporter gene to be activated. The host cell may be any eukaryotic cell, including yeast, zebrafish, C. elegans, Drosophila and mammalian cells having a genome one would like to screen
25   for interactive protein encoding exons or AD encoding exons.

The host cell is constructed to contain and ultimately express a reporter gene having a transcription regulatory
30   element known to include a binding site for the DNA-BD to be employed. The reporter gene product produces a detectable signal when the reporter gene is transcriptionally activated. Thus, a reporter is a moiety whose transcription is detectable, or which expresses a
35   detectable protein or a protein the expression of which may otherwise be determined by monitoring an effect of

expression of the protein.   Examples of reporter gene
products that are readily detectable are well-known and
include: β-galactosidase,   green   fluorescent   protein,
luciferase,   alkaline   phosphatase,   and   chloramphenicol
5   acetyl transferase (CAT) as well as other enzymes and
proteins that are also known as selectable markers.   Other
examples of detectable signals include cell surface markers
such as CD4.   In the exemplified embodiment, the reporter
gene used is the pac gene which encodes the puromycin
10   resistance marker.

In yeast cells, the reporter gene may be homologous
the yeast URA3 gene, the yeast CAN1 gene, the yeast GAL1
gene, the yeast HIS3 gene, or the E. coli LacZ gene.   In
15   mammalian cells, the reporter gene may be homologous to the
CAT gene, the LacZ gene, the SEAP gene, the Luciferase
gene, the GFP gene, the BFP gene, the CD2 gene, the Flu HA
gene, or the tPA gene.

20   The reporter gene in the host cell will be driven by
a transcriptional regulatory element (including promoters
and enhancers) that is capable of binding the DNA-BD
employed in the assay and is functional in the host cell.
Many examples of suitable regulatory elements are well-
25   known, particularly, promoters including those described
below.

The assay may make use of host cells in which the
reporter gene has been previously incorporated, or a
30   construct containing the reporter gene may be introduced to
the cell at the same time as other vectors used in the
assay.

Other vectors used in the assay include a gene trap
35   vector and a two-hybrid vector.   The gene-trap vector is
employed   for   random   insertion   of   a   transcriptional

regulatory protein moiety into the genome of the host cell
and may comprise DNA encoding either a AD or a DNA-BD and
either: an upstream splice acceptor (SA); or, an upstream
transcriptional regulatory element (eg. a promoter) capable
5   of functioning in the host cell for transcription of the
downstream AD or DNA-BD which in turn is followed by an
unpaired splice donor sequence (SD).   In an alternate
embodiment, the gene trap vector has both an upstream SA
and a downstream SD.
10

Incorporation of the gene trap vector within an intron
will permit processing of a chimeric message comprising a
transcript of a flanking endogenous exon joined to the
transcript for the DNA-BD or AD. Use of a gene trap vector
15   having a downstream SD and an upstream promoter is
preferred since transcription of the chimeric message will
not be dependent upon endogenous expression of the host
cell gene.

20   A splice donor (SD) is defined as a nucleotide moiety
having an ability to effect m-RNA splicing to a splice
acceptor site.   Conversely, a splice acceptor (SA) is
defined by its ability to effect mRNA splicing to a splice
donor site. Generally, an unpaired splice donor includes
25   the 3' end of an exon and the 5' end of an intron, and a
splice acceptor includes the 3' end of an intron and the 5'
end of an exon (eg. as defined by Alberts, B. et al.,
at page 373 of Molecular Biology of the Cell (1994), (3d)
Garland Publishing, N.Y. Sequences that may be used as
30   splice acceptors and donors are known and include the
examples of SA and SD sequences as set out in the Examples
herein.

The two-hybrid vector will comprise an upstream
35   transcriptional regulatory element (eg. a promoter) capable
of a functioning in the host cell and driving transcription

- 18 -

of a sequence intended to reconstitute the transcriptional regulatory protein. Thus, the two-hybrid vector will express either a DNA-BD or a AD as the case may be, depending upon the makeup of the gene trap vector.

5   Preferably, the two-hybrid vector will express DNA-BD. The two-hybrid vector also contains a nucleotide sequence which is under the control of the regulatory element and which encodes a selected protein (including a peptide or a polypeptide) of interest (test protein) in respect of which

10 protein-protein interactions are to be determined.

Expression of the two-hybrid vector in the host cell results in the translation of a chimeric protein comprising the transcriptional regulatory protein moiety (eg. DNA-BD)

15 fused with the test protein. Incorporation of the gene trap vector into a gene encoding a protein capable of interaction with the selected protein will result in production in the cell of a reconstituted transcription regulatory protein via interaction of the test protein and

20 the protein product of the trapped gene. Activation of the reporter gene occurs as a result of binding of the DNA-BD to the reporter gene promoter.

Reference herein to "interaction" of proteins (such as

25 an endogenous protein with a test protein) means any interaction whereby proteins tend to be associated in proximity. Such interaction includes any known form of chemical bonding occurring between proteins that are found to be interacting.

30

In an alternate embodiment used for detecting exons encoding endogenous transcription activator domains (protein capable of functioning as an AD), the gene trap vector comprising a DNA-BD is used without a two-hybrid

35 vector. When the gene trap vector integrates into a gene containing an exon that encodes a protein capable of

functioning as an AD in the cell, the resulting gene
product is a chimeric protein that joins both the DNA-BD
coded for by the vector DNA and the AD coded for by the
endogenous exon. Thus, a transcriptional regulatory

5   protein is constituted, capable of activating the reporter
gene in the cell.

A DNA-BD and a AD employed in DNA constructs for use
in this invention may be derived from a single known

10   transcriptional regulatory protein having separate
DNA-binding and transcriptional activation domains (for
example, the yeast GAL4 and GEN4 proteins). Alternatively,
the DNA-BD and AD moieties may be derived from separate
known sources. For example, the DNA-BD may be derived from

15   LexA in E. coli. The DNA-BD may be from DNA binding
proteins other than activators (eg. repressers). The AD
could be derived from amino acids 147-238 of GAL4. The
moieties may also be synthetic, such as the B42 activation
domain. Preferably, the DNA-BD and the AD are from

20   different proteins. In any case, the DNA-BD should not be
capable of functioning significantly as an activator domain
on its own and the AD should not be capable of binding to
the promoter of the reporter gene.

25   In the exemplified embodiment, the DNA-binding domain
is derived from the N-terminal region of the yeast GAL4
protein (eg. amino acids 1-147) and the transcriptional
activation domain is derived from the transcriptional
activator of Herpes Simplex Virus VP16 (eg. amino

30   acids 411-455 of VP16) which is known not bind to DNA but
will function as a transcriptional activator.

The reporter gene may be present in the genome of the
host cell at the time of introduction of the first and/or

35   second DNA constructs. Alternatively, a construct
comprising the reporter gene may be introduced into the

- 20 -

host cell genome at the same time as the first and/or
second DNA constructs. Also, further DNA constructs to be
used in this invention may be introduced to the cell and
made part of the host cell genome before further constructs
5   are introduced, or such constructs may be introduced at the
same time.

DNA constructs, plasmids and the like, as used in this
invention can be delivered or placed in cells *in vivo* using
10   methods known in the art and the methods referred to in the
Examples herein. Such methods include direct injection of
DNA, receptor-mediated DNA uptake or viral-mediated
transfection. Direct injection has been used to introduce
named DNA into cells *in vivo* (see eg. Acsadi *et al.* (1991)
15   Nature 332:815-818; Wolff *et al.* (1990) Science 247:1465-
1468). A delivery apparatus (eg. a "gene gun") for
injecting DNA into cells *in vivo* can be used. Such an
apparatus is commercially available (eg. from BioRad).
Naked DNA can also be introduced into cells by complexing
20   the DNA to a cation, such as polylysine, which is coupled
to a ligand for a cell-surface receptor (see for example
Wu, G. and Wu, C.H. (1998) J. Biol. Chem. 263:14621; Wilson
et al. (1992) J. Biol. Chem. 267:963-967; and U.S. Pat.
No. 5,166,320). Binding of the DNA-ligand complex to the
25   receptor facilitates uptake of the DNA by receptor-mediated
endocytosis. Additionally, a DNA-ligand complex linked to
adenovirus capsids which naturally disrupt endosomes,
thereby releasing material into the cytoplasm can be used
to avoid degradation of the complex by intracellular
30   lysosomes (see for example Curiel *et al.* (1991) Proc. Natl.
Acad. Sci. USA 88:8850; Cristiano *et al.* (1993) Proc Natl.
Acad. Sci. USA 90:2122-2126).

Endogenous genes into which the gene trap vector has
35   integrated may be cloned and sequenced, for example by the
5' RACE method of PCR (polymerase chain reaction).

- 21 -

Furthermore, undifferentiated embryonic stem (ES) cells can be further used to generate mice mutated from the endogenous gene. Heterologous DNA can be inserted into the site of the endogenous gene by known methods including homologous recombination and site directed to recombination.

5' Rapid PCR amplification of cDNA ends (RACE) may be carried out (for example, as described by Skarnes, et al. at (1992) Genes and Development 6, 903-918) to clone a portion of the endogenous gene flanking a gene trap vector insertion. This provides fragments for sequencing and to probe for genes. The source of reagents may be the 5' RACE kit commercially available from Gibco-BRL.

Examples of ES cell lines which may be used in this invention are: porcine (eg. U.S. Patent 5523226 Transgenic Swine Compositions and Methods); murine (eg. D3, R1, CGR8, AB1 ES cell lines); primate (eg. rhesus monkey); rodent; marmoset; avian (eg. chicken); bovine; rabbit; sheep; and horse.

Murine R1 ES cells from A. Nagy [Proc. Nat. Acad. Sci. U.S.A. (1993) 90, 8424-8428] may be grown on Primary Embryonic Fibroblast feeder layers or on gelatinized dishes in the presence of 1000 U/ml murine leukemia inhibitory factor (LIF), ESGRO™ (GIBCO BRL). Selection conditions can be: 150 μg/ml G418, 1.0 μg/ml puromycin, 110 μg/ml Hygromycin B. R1 cells (eg. 2 x 10^7 cells) may be electroporated with, for example, 100 μg linearized DNA in 0.8 ml PBS at 500 μF and 240 V with a BioRad Gene Pulser™ at room temperature.

- 22 -

Example I: Protein Interaction Trap

This aspect of the invention may be conveniently
practiced by modification of standard commercial two-hybrid
5   assay components.   In the following example, the Clontech
Mammalian Matchmaker™ two-hybrid assay kit is modified and
supplemented to provide a reporter gene as a selectable
marker (pac) for puromycin resistance; a DNA-BD from GAL4
(as provided in the commercial kit); and an AD from Herpes
10   Simplex Virus VP16 (as provided in the kit).   In this
example, all DNA constructs, including the reporter gene
are introduced into a murine R1 ES cell line host cell.

The first DNA construct (two-hybrid vector) comprises
15   a sequence encoding a GAL4 DNA-BD which recognizes a
binding site on the reporter gene and further comprises a
sequence encoding p53 protein (Clontech, pM-53 plasmid).

The second DNA construct (gene trap vector) is novel
20   and comprises a promoter capable of operation in the host
cell, driving a VP16 AD upstream of a splice donor
sequence.   In an alternate embodiment, the novel gene trap
vector does not contain a promoter and has a splice
acceptor sequence upstream of the VP16 AD followed by a
25   poly-adenylation signal.

When the gene trap is integrated into an intron
adjacent to an exon of the host cell encoding a protein
domain capable of interaction with p53 protein, a
30   transcriptional regulatory protein comprising GAL4 BD and
the VP16 AD is constituted.   Expression of the reporter
gene in a host cell as a result of binding by the DNA-BD is
detected by culturing the transformed cells in the presence
of puromycin.   Cells in which the reporter gene has been
35   activated will survive.   Alternatively, the reporter used
in the assay could remain as CAT and determination of

- 23 -

reporter gene activity may be carried out according to
standard assay procedures, for example as taught in the
Clontech kit instructions.

5   Host cells are transformed by any of the well-known
methods, selected as being suitable for the particular cell
type.   Electroporation or calcium phosphate mediated
transfection   are   suitable   for   mammalian   cells.
Transfection procedures as taught in the Clontech kit
10   instructions may be used. A preferred method known for ES
cells is electroporation.

The following plasmids are constructed and/or employed
in this example.   The reporter (pG5Puro) is a modified
15   version of the GAL4 responsive CAT reporter construct from
the Clontech Matchmaker™ kit (pG5CAT).   In this example,
the CAT reporter gene is replaced by the selectable marker
pac, generating a reporter construct containing the
puromycin resistance gene under the control of the
20   adenovirus E1b minimal promoter used in the Clontech
plasmid.   Upstream, are five copies of the 17 nucleotide
consensus GAL4 binding site (galactose upstream activating
sequence: UAS$_G$).

25   The second plasmid is the pM-53 vector from the
Matchmaker™ kit which is an expression plasmid containing
the SV40 promoter driving a GAL4 DNA-BD.   The commercial
construct encodes p53 protein, but the multiple cloning
site downstream from the DNA-BD may be used to insert
30   different bait proteins. This functions as the two-hybrid
vector.

A gene trap vector plasmid is constructed by inserting
an oligomer sequence encoding a consensus SD sequence in
35   frame into a SalI/BspMI digested pVP16 plasmid (Clontech)
simultaneously   deleting   the   stop   codons   and

- 24 -

poly-adenylation signal. Thus, a gene trap vector is
generated comprising an SV40 promoter driving expression of
the AD.   Three versions of this vector were created
resulting in splicing in all three potential reading
5   frames.   The following are examples of consensus SD
sequences:

            AGGTAAGT (SEQ ID NO:1)
            AGGTGAGT (SEQ ID NO:2)
10

, each of which may be preceded by C or A.

            An alternate gene trap vector plasmid may be
constructed containing the VP16 AD downstream of a SA
15   sequence.   Three constructs should be generated, each
resulting in splicing in each of three possible reading
frames.   SA sequences comprise a polypyrimidine tract
followed by a nucleotide, T or C, AG, and at least G or A.
Examples are the murine En-2 splice acceptor and the splice
20   acceptors from human β-globin and rabbit b-globulin.

            The following methods may be used for construction of
VP16 gene trap vectors:

25   (I)   To construct the gene trap vector consisting of
                  the SV40 promoter driving the expression of VP16
                  fused to an unpaired splice donor sequence:

                  (a)   Digest pVP16 (Clontech) with SalI and BspMI;
30

                  (b)   Isolate and purify the 3.0 kb fragment;

                  (c)   Ligate the 3.0 kb pVP16 fragment with each
                        of the following pairs of oligomers to
35   create fusions of VP16 with unpaired splice

- 25 -

donor sequences in all three possible reading frames:

Pair #1:   5' tcgacaggtaagt 3' (SEQ ID NO:3)
            5' tcatacttacctg 3' (SEQ ID NO:4)

Pair #2   5' tcgaccaggtaagt 3' (SEQ ID NO:5)
            5' tcatacttacctgg 3' (SEQ ID NO:6)

Pair #3   5' tcgacccaggtaagt 3'(SEQ ID NO:7)
            5' tcatacttacctggg 3'(SEQ ID NO:8)

(II) To construct an alternate gene trap vector comprising the En-2 SA sequence fused 5' of the VP16 transcriptional activator:

(A) (1)   digest pGT4SA vector (Gossler et al. 1989 Science 244:463-465) with Xbal;

     (2)   fill in ends with T4 DNA polymerase to generate blunt ends;

     (3)   digest with NdeI; and

     (4)   Isolated and purify the 2.0 kb fragment encoding the En-2 splice acceptor sequence.

(B) (1)   digest pVP16 (Clontech) with NheI;

     (2)   fill in ends with T4 DNA polymerase to generate blunt end;

     (3)   digest with NdeI; and

     (4)   isolated and purify the 2.8 kb fragment encoding the VP16 transcriptional activator sequence.

- 26 -

(C)   Ligate 2.0 kb En-2 splice acceptor fragment
      to 2.8 kb VP16 containing vector.

(D)   To generate SA-VP16 in the other two
      potential reading frames:

      (1)   digest the above vector with SexAI and
            BglII;

      (2)   ligate the following pairs of oligomers
            to generate fusions in the other two
            possible reading frames:

            Pair #1   5' ccaggtcgca 3' (SEQ ID NO:9)
                       5' gatctgcga 3' (SEQ ID NO:10)

            Pair #2   5' ccaggtgca 3' (SEQ ID NO:11)
                       5' gatctgca 3' (SEQ ID NO:12)

The three forms of the gene trap vector representing
all three potential reading frames are placed in a head to
tail tandem array allowing the use of alternate promoters
to generate three hybrid mRNAs fusing the VP16 domain in
all three possible reading frames to a adjacent exon upon
integration into a gene within the host cell genome.

The following protocol may be followed:

1.   Construct a reporter murine embryonic stem (ES) cell
line using standard methods by co-electroporation of
linearized pG5Puro, pM-53 and pPGKHyg into the murine R1 ES
cell line.   Hygromycin resistance is used to monitor
transfection efficiency.

2.   Characterize the reporter cell lines for its ability
to detect protein-protein interactions by electroporating
with pVP16T (Clontech) as a positive control and pVP16-CP

- 27 -

(Clontech) as a negative control for protein-protein interaction. pVP16T expresses a fusion of the VP16 activation domain to the SV40 large T antigen, which is known to interact with p53. The pVP16-CP negative control
5   plasmid expresses a fusion of the VP16 activation domain to a viral coat protein, which does not interact with p53.

3.   Upon electroporation of positive or negative control plasmids, cells are then placed under 1.0 ug/ml puromycin
10   selection.

4.   Select appropriate reporter cell clones that confer puromycin resistance in the presence of VP16T but not with pVP16-CP (cells express pG5Puro and pM-53).
15

5.   Electroporate gene trap vectors into reporter cell line and select for puromycin resistance with 1.0 ug/ml puromycin.

20   6.   Pick individual puromycin resistant colonies and isolate RNA from each clone.

7.   Isolate and sequence trapped exon/gene by rapid amplification of cDNA end (RACE) PCR (eg. see: Skarnes,
25   et al. 1992. *Genes and Development* 6:903-18). Clontech sequencing primers for VP16 may be used.

Example II: Transcriptional Activator Domain Trap

30   In this example, the methods employed in the preceding example are used in an assay employing the ES host cell, the same reporter gene construct (pG5CAT) employed in the preceding example, and a gene trap vector plasmid designed to trap genes expressing endogenous protein capable of
35   functioning as a transcriptional activator domain (AD) in conjunction with the DNA-BD expressed by the gene trap

- 28 -

vector.   Expression of chimeric proteins comprising the
DNA-BD   fused   to   an   endogenous   protein   capable   of
functioning   as   a   AD   will   result   in   activation   of   the
reporter   gene   which   comprises   a   binding   site   for   the
5   DNA-BD.

The   gene   trap   vector   plasmid   is   constructed   by
inserting   an   oligomer   sequence   encoding   the   consensus   SD
sequence   in   frame   into   the   SalI/BspMI   digested   pM   plasmid
10   (Clontech)   resulting   a   vector   comprising   of   the   SV40
promoter   driving   the   GAL4   DNA-binding   domain   linked   to   a   SD
sequence.   Three   versions   of   this   vector   are   created
resulting   in   splicing   in   each   of   the   three   potential
reading   frames,   respectively.   A   consensus   splice   donor
15   sequence   domain   contains   the   following:

Exon....AGGTAAGT...Intron (SEQ ID NO:1)

To   construct   the   vector   consisting   of   the   SV40
20   promoter   driving   the   expression   of   the   GAL4   DNA   binding
domain   fused   to   an   unpaired   splice   donor   sequence:

(a)   digest pM (Clontech) with SalI and BspMI;

25   (b)   isolate and purify the 3.2kb fragment; and

(c)   ligate the 3.2 kb pVP16 fragment with each of the
following pairs of oligomers to create fusions of
VP16   with   SD   sequences   in   all   three   possible
30   reading frames:

Pair #1   5' tcgacaggtaagt 3' (SEQ ID NO:3)
5' tcatacttacctg 3' (SEQ ID NO:4)

35   Pair #2   5' tcgaccaggtaagt 3'(SEQ ID NO:5)
5' tcatacttacctgg 3'(SEQ ID NO:6)

- 29 -

Pair #3   5' tcgacccaggtaagt 3' (SEQ ID NO:7)
           5' tcatacttacctggg 3' (SEQ ID NO:8)

The three forms of the gene trap are then placed in a
5 head-to-tail tandem array allowing the use of alternative
promoters to generate three hybrid mRNAs fusing the GAL4
DNA domain in all three possible reading frames to the next
endogenous exon upon integration into a gene within the
genome.
10

The following protocol may be used:

1.   Construct a reporter murine embryonic stem (ES) cell
line using standard methods by co-electroporation of
15 linearized pG5Puro, and pPGKHyg into the murine R1 ED cell
line.

2.   Select, and expand several clones which contain
pG5Puro.
20

3.   Characterize the reporter cell line for ability to
express   transcriptional   activator   domains   by
electroporating with pM3-VP16 (Clontech) as a positive
control and pM-53 (Clontech) as a negative control for
25 transcriptional activator domains.   pM3-VP16 expresses a
fusion of the VP16 activation domain to the GAL4 DNA
binding domain which is known transactivate the GAL4
responsive promoter in pG5Puro.   The pm-53 negative control
plasmid expresses a fusion of the VP16 activation domain to
30 p53, which does not transactivate the GAL4 responsive
promoter in pG5Puro.

4.   Upon electroporation of positive or negative control
plasmids, cells are then placed under 1.0 ug/ml puromycin
35 selection.

5.   Select appropriate reporter cell clones that confer puromycin resistance in the presence of pM3-VP16 but not with pM-53.

6.   Electroporate gene trap vector into reporter cell line and select puromycin resistance with 1.0 ug/ml puromycin.

7.   Pick individual puromycin resistant colonies and isolate RNA from each clone.

8.   Isolate and sequence trapped exon/gene by rapid amplification of cDNA ends (RACE-PCR).

All publications and patents cited in this specification are incorporated herein by reference. Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to those of ordinary skill in the art in light of the teachings of this invention that changes and modification may be made thereto without departing from the spirit or scope of the appended claims.

WE CLAIM:

1. A method for detecting interaction between an endogenous protein of a cell and a test protein, wherein said cell contains a first DNA sequence encoding a reporter under transcriptional control of a transcriptional regulatory element, and a second DNA sequence that is expressed by the cell and which encodes a first hybrid protein comprising:

    (a) a first transcriptional regulatory protein moiety selected from the group consisting of: a DNA-BD that recognizes a binding site on the transcriptional regulatory element controlling transcription of the first DNA sequence and, a AD functional in the cell; and

    (b) a test protein;

wherein the method comprises the steps of:

    (a) placing into the cell or an ancestor of the cell, a DNA construct comprising one or more m-RNA splice sites, and a third DNA sequence encoding a second transcriptional regulatory protein moiety which, when combined with the first transcriptional regulatory protein moiety will reconstitute a transcriptional regulatory protein capable of binding to and activating the transcriptional regulatory element controlling transcription of the first DNA sequence; and,

    (b) determining whether the reporter is expressed by the cell or a descendant of the cell, as an indicator of expression of a second hybrid protein comprising the second transcriptional

regulatory protein moiety and an endogenous
protein of the cell capable of interaction with
the test protein.

5   2.   The method of claim 1 wherein the DNA construct
comprises the third DNA sequence upstream of a SD.

3.   The method of claim 2 wherein a transcriptional
regulatory element is operably linked to the third DNA

10   sequence.

4.   The method of claim 1 wherein the DNA construct
comprises a SA upstream from the third DNA sequence and
does not comprise a transcriptional regulatory element.

15

5.   The method of claim 4 wherein the DNA construct
comprises a poly-adenylation signal downstream from the
third DNA sequence.

20   6.   The method of claim 1 wherein the DNA construct
comprises the third DNA sequence, an upstream SA, and a
downstream SD.

7.   The method of any one of claims 1-6 wherein the third

25   DNA sequence encodes an AD.

8.   The method of any one of claims 1-7 wherein the DNA
construct encodes only a transcriptional regulatory protein
moiety between a first position on the construct defined as

30   a 5' end of the third DNA sequence or a SA, and a second
position on the construct defined as a SD or a
poly-adenylation signal.

9.   A method for detecting an endogenous transcription

35   activator domain (AD) of a cell, wherein the cell contains
a   first   DNA   sequence   encoding   a   reporter   under

transcriptional control of a transcriptional regulatory element, wherein the method comprises the steps of:

    (a)   placing into the cell or an ancestor of the cell, a DNA construct comprising a m-RNA splice site and a second DNA sequence encoding a DNA-BD that recognizes a binding site on the transcriptional regulatory element controlling transcription of the first DNA sequence; and

    (b)   detecting expression of the reporter in the cell or a descendant of the cell, as an indicator of expression of a hybrid protein comprising the DNA-BD and an endogenous protein of the cell capable of functioning as an activator domain.

10. The method of claim 9 wherein the DNA construct comprises the second DNA sequence upstream of a SD.

11. The method of claim 10 wherein a transcriptional regulatory element is operably linked to the second DNA sequence.

12. The method of claim 9 wherein the DNA construct comprises a SA upstream from the second DNA sequence and no transcriptional regulatory element.

13. The method of claim 12 wherein the DNA construct comprises a poly-adenylation sequence downstream from the second DNA sequence.

14. The method of claim 9 wherein the DNA construct comprises the second DNA sequence, with an upstream SA and a downstream SD.

15. The method of any one of claims 9-14 wherein the DNA
construct encodes only a transcriptional regulatory protein
moiety between a first position on the construct defined as
a 5' end of the second DNA sequence or a SA, and a second
5   position   on   the   construct   defined   as   a   SD   or   a
poly-adenylation signal.

16. A DNA construct as defined in claim 8.

10   17. A DNA construct as defined in claim 15.

18. A cell as defined in claim 1 or 9.

19. The cell of claim 18 transformed with a DNA construct
15   of claim 16 or 17.

20. A method of making an array of DNA constructs of
claims 16 or 17, comprising the steps of joining a DNA
sequence encoding a transcriptional regulatory protein
20   moiety in each of three possible reading frames with a DNA
sequence encoding a splice acceptor (SA) or a splice donor
(SD).

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6   C12Q1/68   C12N15/10   C12N15/85   C12N15/62

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6   C12Q   C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	FIELDS S ET AL: "A NOVEL GENETIC SYSTEM TO DETECT PROTEIN-PROTEIN INTERACTIONS" NATURE, vol. 340, 20 July 1989, page 245/246 XP002914166 cited in the application see the whole document ---	18
A	WO 97 30164 A (MEDICAL RES COUNCIL ;HASELOFF JAMES PHILLIP (GB); HODGE SARAH (GB)) 21 August 1997 see claims --- -/--	16,17

[X] Further documents are listed in the continuation of box C.   [X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 June 1999	30/06/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Molina Galan, E

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	SKARNES W C: "THE IDENTIFICATION OF NEW GENES: GENE TRAPPING IN TRANSGENIC MICE" CURRENT OPINION IN BIOTECHNOLOGY, vol. 4, 1 January 1993, pages 684-689, XP000569473 cited in the application	
A	WO 95 26400 A (ONYX PHARMA INC) 5 October 1995 cited in the application	
A	WO 94 16092 A (JARVIK JONATHAN WALLACE) 21 July 1994 cited in the application	

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 9730164	A	21-08-1997	AU	1801497 A	02-09-1997
			CA	2246414 A	21-08-1997
			EP	0880593 A	02-12-1998
WO 9526400	A	05-10-1995	US	5525490 A	11-06-1996
			AU	2779095 A	17-10-1995
WO 9416092	A	21-07-1994	US	5652128 A	29-07-1997
			EP	0656954 A	14-06-1995